NEW TRAINING ALGORITHM FOR NEURAL NETWORKS

Joseph C. Chow, Lalita Udpa and Satish Udpa

Department of Electrical and Computer Engineering and
Center for NDE
Iowa State University
Ames, Iowa 50011

INTRODUCTION

Neural networks are finding increasing use as an adaptive signal classifier in many engineering applications. Artificial neural networks have been used for classifying NDE signals such as ultrasonic and eddy current signals. These networks consist of densely interconnected units with variable interconnection weights. The networks can be categorized according to their architecture and learning algorithm. The class of neural networks most commonly used is the multilayer perceptron network. The basic structure of this network consists of one input layer, one output layer, and one or more hidden layers. In order to properly classify input signals, the neural network must first be trained. The network is trained by presenting known input patterns to the input nodes and the corresponding desired output patterns to the output nodes. The network computes the output corresponding to each input and the error between the desired output and the actual output is used to drive an iterative algorithm for updating the weights in a manner that minimizes the error.

Conventionally, the error is minimized using the gradient descent method. However, there are many disadvantages associated with this method. For example, the solution tracked using this method can easily converge to local minima. In addition, it is also possible that the solution will never converge and will oscillate between multiple points. In fact, these difficulties are encountered in nearly all high dimensional, nonlinear optimization problems.

This paper presents an innovative approach for training multilayer perceptron networks using the homotopy continuation method [4] [5] [6]. Continuation methods are used extensively for solving optimization problems in many fields such as engineering, mathematical analysis, and most recently, signal processing [1] due to their ability to provide globally convergent solutions. In this approach, the homotopy function defines a family of paths from solutions of a simple problem to the solutions of the more complex problem. The proposed technique offers the globally optimum solution which usually translates into better classification performance.

HOMOTOPY CONTINUATION METHODS

Homotopy continuation methods are numerical techniques used to determine the zeroes of a system of nonlinear equations [2][3]. The underlying concept of the method involves identifying a simple system with known solution and slowly deforming it into the desired system with unknown solutions. During the deformation process, a family of paths is defined from the solutions of the known system to the solutions of the unknown system.

The continuation method is a numerical procedure used to track this solution path, where every point on the path represents the solution of the deformed system, for a particular value of the deformation parameter. When the known system is completely deformed into the unknown system, the desired solutions to the unknown system are obtained. Unlike the gradient based numerical methods, such as the steepest descent methods, homotopy continuation methods are globally convergent and solution exhaustive. Globally convergent implies that the method will converge to a solution irrespective of the location of the initial starting point. Solution exhaustive signifies that all solutions to a desired system can be found. However, several conditions must be met for the solution exhaustive property to hold. First, the system of interest must consist of polynomial equations of known degree. Second, the initial system must also consist of polynomial equations with the same degree as the unknown system.

The homotopy function h(x,t) is defined as:

$$h(x,t) \equiv (1 - t)\, g(x) + t\, f(x) \tag{1}$$

where t is the homotopic tracking parameter, g(x) is the system with known solutions and f(x) is the system of interest whose solutions are desired. As the parameter t varies from 0 to 1, h(x,t) deforms the known system, g(x), to the the desired system, f(x). For example, at $t = 0$, $h(x,0) = g(x) = 0$ gives the initial starting solutions of the known system.When $t = 1$, $h(x,1) = f(x) = 0$ gives the desired solution of the system of interest. The intermediate values of t between 0 and 1 correspond to the different deformed functions, h(x,t).

As t varies from 0 to 1, a family of paths is tracked from the solutions of $g(x) = 0$ to the solutions of $f(x) = 0$. A typical continuation path for a one dimensional function is illustrated in Figure 1. The continuation path is tracked by solving $h(x,t) = 0$ at each incremental step of t. Since every point (x,t) on the path satisfies the equation $h(x,t) = 0$, the derivative of the homotopy function with respect to the tracking parameter, t, is equal to zero.The homotopy differential equation is obtained by differentiating Eq. (1) with respect to the tracking parameter,

$$\frac{d}{dt}h(x(t)) = H_x(x(t))\frac{dx(t)}{dt} = 0 \tag{2}$$

where H_x is the Jacobian matrix of the homotopy function with respect to x. Eq. (2) can be rewritten as

$$\frac{dx(t)}{dt} = - \left[H_x(x(t))\right]^{-1} \frac{\partial h(x(t))}{\partial t} \tag{3}$$

where Eq. (3) is referred as the homotopy differential equation. The path is guaranteed to be continuously differentiable if the Jacobian matrix $H_x(x(t))$ of the homotopy function is of full rank. While Eq. (3) offers a precise description of the solution path, it has limitations when path types as illustrated in Figure 2 are encountered. This path demonstrates cases where the variation of t is not isomorphic with x. This problem is overcome by parameterizing the path in terms of its arc length, s.

It can be shown that the homotopy equation using s as the tracking parameter is

$$\left[\frac{dx(s)}{ds}\right]_{-i} = - \left[H_{x,t}(x(s))\right]^{-1}_{-i} \frac{\partial h(x(s))}{\partial x_i} \tag{4}$$

where A_{-i} denotes the matrix A with its $i^{\underline{th}}$ column removed. Solving the homotopy differential equation provides a path (x(s),t(s)), where by changing s, one can track the solution for the variation of t from 0 to 1.

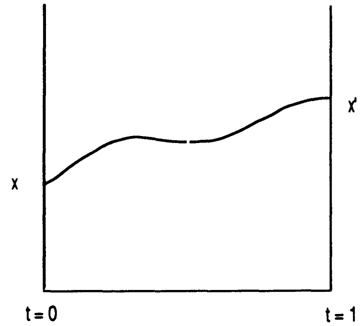


Figure 1 Motononic homotopy path

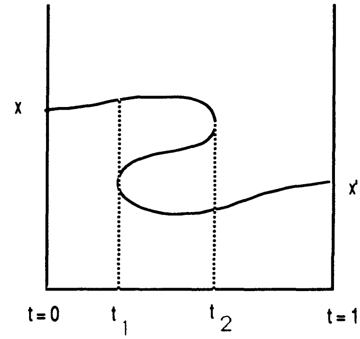


Figure 2 Nonmonotonic homotopy path

There are several variations of the homotopy functions as depicted in the following equations.

$$h(x,t) = (1-t)\, g(x) + t\, f(x) \tag{5}$$

$$h(x,t) = (1-t)\, (x-x_o) + t\, f(x) \tag{6}$$

$$h(x,t) = f(x) - (1-t)\, f(x_o) \tag{7}$$

Eq. (5) is known as the all-solution homotopy method. In this approach, if g(x) and f(x) are nonlinear systems of the same degree, then an exhaustive solution set of the unknown system can be obtained. This method is most appropriate for solving systems which consist of polynomial functions since the degree of a polynomial system can be easily determined. Eq. (6) and (7) are known as the fixed-point homotopy method and the Newton homotopy method, respectively. Both methods entail tracking only one solution at a time and are globally convergent.

BACKWARD ERROR PROPAGATION ALGORITHM

The backward error propagation algorithm, commonly used in the training of multilayer perceptron networks, simply propagates the error between the desired output and the actual output back through the network to adjust the interconnection weights. The error function is expressed as

$$E = \frac{1}{2} \sum_{c} \sum_{j} (y_j(c) - d_j(c))^2 \tag{8}$$

where
c is the input sample case index
j is the output node index
y is the actual output
d is the desired output

The overall objective is to minimize the error function by adjusting the interconnection weights. Using the notation shown in Figure 3, the minimization equations for the output layer connections are

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial x_j} \frac{\partial x_j}{\partial w_{ij}} \tag{9}$$

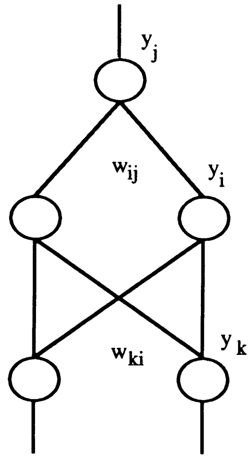


Figure 3 Simplified multilayer perceptron network

$$\frac{\partial E}{\partial \theta_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial \theta_j} = 0 \tag{10}$$

where the nodal activation function is given by the sigmoidal function

$$y_j = f(x_j) = \frac{1}{1 + e^{-(x_j+\theta_j)}} \tag{11}$$

and

$$x_j = \sum_i w_{ij}\, y_i \tag{12}$$

Similarly, for the hidden layer connections, the minimization equations are

$$\frac{\partial E}{\partial w_{ki}} = \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial x_i}\frac{\partial x_i}{\partial w_{ki}} \tag{13}$$

$$\frac{\partial E}{\partial \theta_i} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial x_j}\frac{\partial x_j}{\partial y_i}\frac{\partial y_i}{\partial \theta_i} = 0 \tag{14}$$

The learning algorithm solves for w and θ by minimizing the error function, E, iteratively using the gradient descent method. Two distinct disadvantages associated with gradient methods are, as mentioned before, convergence to local minima and failure to converge.

NEURAL NETWORK TRAINING USING HOMOTOPY CONTINUATION METHODS

In order to obtain the optimum solution for the network, an exhaustive solution set must first be obtained. The all-solution homotopy method allows all solutions to be tracked provided the system of interest consists of polynomial functions. Consequently, the error minimization equations must be formulated in terms of polynomials. This was done by approximating the sigmoidal function using polynomials. The third order polynomial approximation was determined as

$$p(x) = 0.5 + 0.19745\,x + 2.5224 \times 10^{-8}\,x^2 - 4.3917 \times 10^{-3}\,x^3 \tag{15}$$

$$-3.8 \leq x \leq 3.8$$

where the bounded domain is chosen [-3.8, 3.8]. The solutions of this system, using the all-solution method, are obtained by tracking a simpler system of equations with known solutions. However, the problem of generating a high order initial system with known real solutions is still unresolved. An attempt was made to construct this known system using randomly selected real numbers as the solutions. The method involves the solution of an over-determined system of equations. The results using this method were unsatisfactory since the matrix equations representing the over-determined system is ill-conditioned. Consequently, the estimates obtained have large variance. The generation of a large system of nonlinear polynomial functions is a major challenge and is currently an area of active research. As a result, the problem was solved using the fixed-point homotopy function. Fixed-point homotopy is frequently used due to its linear nature and the global convergence property which allow it to track a solution of nonlinear systems.

Let $\mathbf{x} = (\theta,\mathbf{w})$ be the desired solution vector consisting of the weights $\mathbf{w}$ and biases, θ. Applying the fixed point homotopy method, the equation for the problem is

$$\mathbf{h}(\mathbf{x},t) = (1-t)\,(\mathbf{x}-\mathbf{x}_0) + t\,\mathbf{f}(\mathbf{x}) \tag{16}$$

where $\mathbf{x}_0$ is an initial point and $\mathbf{f}(\mathbf{x})$ is obtained from Eq. (9), (10), (13), and (14).

SIMULATION RESULTS

The architecture used to demonstrate the fixed-point method is shown in Figure 4. The network consists of two input nodes, one hidden node, and one output node. Two bias variables are used to enhance the output convergence. Four two-class training patterns were considered. Two of the examples are linearly separable whereas the other examples involve a nonlinear boundary, as illustrated in Figure 5.

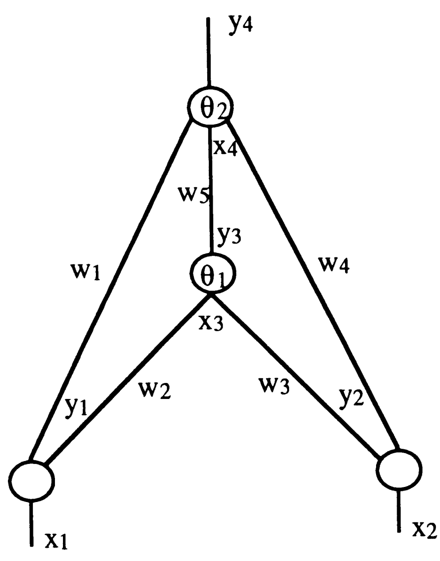
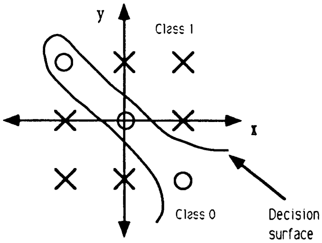
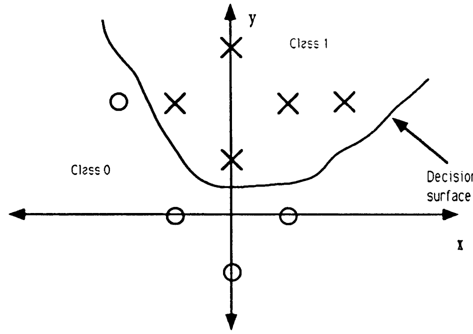


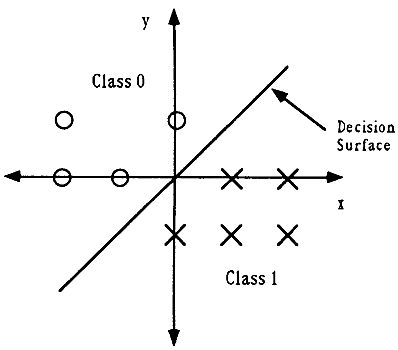
Figure 4 Experimental network



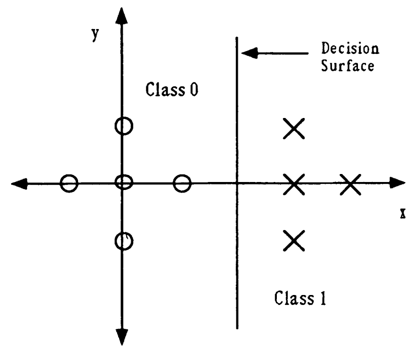
Two class input sample problem #1



Two class input sample problem #2



Two class input sample problem #3



Two class input sample problem #4

Figure 5 Two class training patterns

The results of implementing the homotopy method and the corresponding results of the gradient method are summarized in Table 1. In the case of linearly separable problems, both methods converged to the desired solution for most initial values. However, for larger values of initial weights, the gradient method fails to converge, whereas the fixed-point homotopy method still converges to the desired solution. For the nonlinearly separable cases, fixed-point method converges to the desired solution for most initial points whereas the gradient method fails more frequently. In addition, there are instances where fixed-point method failed to converge to the desired solution. This can be attributed to the fact that the fixed point homotopy method is not solution exhaustive.

Table 1. Summary of Classification Performance

Pattern Number	Homotopy Method	Gradient Method
1	100%	56%
2	100%	11%
3	100%	56%
4	100%	100%

CONCLUSION

The initial results of implementation are very encouraging. The homotopy training method ensures convergence of the algorithm in relatively fewer iterations resulting in lesser training time for the neural network. The fixed point homotopy method does not offer an exhaustive set of solutions. However, it does provide solutions in cases where the gradient method failed. However, a major challenge in the application of the homotopy continuation method is the generation of the system of polynomial equation with known solutions. The direction for future work includes the development of a suitable technique for generating exhaustive solutions of the system which will result in the determination of the global minimum of the error surface.

REFERENCES

1. V. L. Stonick, "Global Methods of Pole/Zero Modeling for Digital Signal Processing using Homotopy Continuation Methods," Ph.D. Dissertation, North Carolina State University, 1989

2. C. B. Garcia and W. I. Zangwill, Pathways to Solutions, Fixed Points and Equilibria, Prentice-Hall, 1981

3. A. P. Morgan, Solving Polynomial Systems Using Continuation for Engineering and Scientific Computations, Prentice-Hall, 1987

4. J. Chow, L. Udpa and S. Udpa, "Homotopy Continuation Methods for Neural Networks," ISCAS Conference, Singapore, June 1991

5. J. Chow, L. Udpa and S. Udpa, "Homotopy Continuation Method for Neural Networks", To Appear in IEE Second International Conference and Exhibition on Artificial Neural Networks, London, England, Nov. 1991

6. J. Chow, L. Udpa and S. Udpa, "Neural Network Training Using Homotopy Continuation Methods," to appear in IJCNN, Singapore, Nov. 1991